

# LAV Project

## Report

**Elijah Relf** : s3564403

**Vinesh Singh Gobin** : s3670123

**Rohit Menon** : s3544222

# Table of Contents

---

Introduction	<b>3</b>
The Data	<b>3</b>
Hypothesis Questions	<b>5</b>
The Cleaning Process	<b>5</b>
The Wrangling Process	<b>8</b>
The Sampling Process	<b>11</b>
Overview of Sampling & Road to Visualisation	<b>16</b>
Visualisation	<b>17</b>
Conclusions	<b>37</b>
References	<b>38</b>

# Introduction

The purpose of this investigation is to analyse the **On-Street Car Parking Sensor Data from 2017** (<https://data.melbourne.vic.gov.au/Transport/On-street-Car-Parking-Sensor-Data-2017/u9sa-j86i>) provided by the City of Melbourne for our sponsor.

## The Data

The data set is provided by the City of Melbourne, the data is gathered by in-ground parking bay sensors in a variety of areas in and around the CBD. The sensors record when a vehicle arrives and when it departs and includes information such as:

- Parking restrictions for the bay
- Whether the vehicle has overstayed the restriction (has been issued a fine)

Also, there are 3 known data issues that have been outlined by City of Melbourne. These are:

- There are 860 records that have the text 'OLD' at the end of the restriction in the field named 'Sign'. This has been investigated and determined the text refers to one of two things, either:
  - At the time of the parking event the restriction that has been captured is correct but when the data was extracted from the server (April 2018) the restriction had since changed.
  - The sensor has since been replaced.
- There are a number of records that have a negative value in the duration in field 'seconds'. This occurs due to a sensor detecting an arrival time being after the departure time. .

# The Data (Continued)

- There are several records where the arrival and departure times have not been recorded. If the time has not been recorded the time is calculated from midnight of the arrival day to either midnight of the departure day or to the departure time

The data set contains roughly 35.9million rows with 14 columns, the columns are the following:

- Device ID – the id of the in-ground sensor
- Arrival Time – Date & Time that the sensor detected a vehicle located in the bay
- Departure Time – Date & Time that the sensor detected a vehicle no longer in the bay
- Duration Seconds – The time difference between the arrival time & departure time
- Street Marker – located next to the parking bay with a unique id
- Sign – The parking sign currently in effect at the time of the parking event
- Area – the area the parking bay is in
- Street ID – a GIS key that describes the street segment where the sensor is located
- Street Name – Where the bay / park is located
- Between Street 1 – Closest intersecting street – ideally the next one in front of the parked vehicle
- Between Street 2 – Closest intersecting street – ideally the one behind the parked vehicle
- Side of Street – Side of street of which the parking event occurred  
1 – Centre    2 – North    3 – East    4 – South    5 – West
- In Violation – whether the vehicle was in violation
- Vehicle Present – was the vehicle present for the violation

# Hypothesis Questions:

Before we started looking at the data, we came up with possible questions that we may be able to answer by looking at the column names.

## What will the data be used for:

Which areas are the busiest?( to increase/decrease size)

Which parking spots are used and the duration (Full 2 hours or violation):  
Hours could be decreased so more people could use it. Also to increase the size?

They could use the data to see if a carpark needs to be removed or is not being used.

See which ones get fined the most – To send more ticket inspectors there.

What time do the parking spots get filled up the most, and when are they most empty. –Ticket inspectors can be sent there 2 hours after peak to check who has violate?

Is there a relation between parking spots and traffic on the street?

Street Markers can be used to determine which parking spots on a street are the most popular.

## The Cleaning Process

Due to the large quantity of data, we decided instead of loading the complete data set into Jupyter to split the files so that each of us can work on one individually. This was made possible by the **splitting script for csv.ps1** which we found online and altered to our needs (<https://www.codetwo.com/admins-blog/how-to-split-csv-file-into-multiple-files-using-powershell/>). The split files were created by splitting the file into thirds with each csv containing around 11.9 million rows. From this point we would then load the csv into Jupyter to start delving into the data.

# The Cleaning Process (Continued)

We looked at locating the most common problems with data, such as null values, sanity issues and spacing errors. For example:

- Null Errors

```
1 df1['DepartureTime'].isnull().sum()
```

- Sanity Checks

```
for data in df1['Side Of Street']:
    if data < 1 or data > 5:
        print(data)
```

- Spacing Errors

```
df1['Area'] = df1['Area'].str.strip()
```

Upon each individual's completion of taking a look at their data sets, we found that the only issues were that the sign column had many null values, after attempting to fix this issue we found it to be too complex to do in a short amount of time. Thus, we decided we should just drop the rows with any NaN values – which was about a 34% loss in data. Though this still leaves us with roughly 22 million rows of data to analyse.

Due to the similarities in everyone's data sets, we decided to create a program that cleaned the main CSV file instead of them each.

The **datapreprocessing.py** program was created to do this for us, it uses the pandas library and the system module to assist us in cleaning.

# The Cleaning Process (Continued)

```
def main(csvfile):
    chunkies = pd.read_csv(csvfile, sep=',', chunksize=250000, low_memory=False)
    write_header = True
    chunkno = 1
    for chunk in chunkies:
        print('Cleaning Chunk: {}'.format(chunkno))
        process(chunk)
        chunk.to_csv('cleaned.csv', mode='a', header=write_header, index=False)
        write_header = False
        print('Successfully Appended to Cleaned CSV File')
        print('-----')
        print()
        chunkno += 1
    print('Done.')
    sys.exit()

if __name__ == "__main__":
    csv = input('Enter CSV File: ')
    main(csv)
```

The program takes in the CSV as input, for which it then starts looping through chunks of the data (each chunk containing 250,000 rows) that gets cleaned by the process function

```
def process(chunk):
    print('-----')
    print('Dropping NaN Values')
    chunk.dropna(axis=0, how='any', inplace=True)
    print('Chunk Drop Done')
    print('-----')
    print()
    print('-----DateTime Alterations-----')
    print('Converting Seconds to Datetime')
    chunk['DurationSeconds'] = chunk['DurationSeconds'].astype('float64')
    chunk['DurationSeconds'] = pd.to_datetime(chunk['DurationSeconds'], unit='s')
    chunk['DurationSeconds'] = chunk['DurationSeconds'].apply(lambda x:x.time().strftime('%H:%M:%S'))
    print('Seconds Conversion Completed...')
    print('Converting ArrivalTime to Datetime')
    chunk['ArrivalTime'] = pd.to_datetime(chunk['ArrivalTime'])
    print('ArrivalTime Successfully Converted...')
    print('Departure Time Conversion')
    chunk['DepartureTime'] = pd.to_datetime(chunk['DepartureTime'])
    print('DATE TIME CONVERSIONS COMPLETED')
    print('-----')
    print()
```

# The Cleaning Process (Continued)

For every chunk in the CSV, process drops all rows with NaN / Null values in a column, converts Duration Seconds to a datetime object – which then gets converted to its necessary HH:MM:SS format, converts Arrival & Departure Time to datetime format and strips any necessary empty space from the relevant columns.

After each chunk has been cleaned, it is then appended to the cleaned.csv.

## The Wrangling Process

To gain a greater understanding of the data we possess, we split up the cleaned.csv into 3 separate CSV files with roughly 8 million rows in each set. The way the data was split translates to the beginning, the middle and the end of the data set. We did this so that when we visualise them for our sampling, we may be able to see if the data holds any disparities or if all of it roughly looks the same.

Our main thought behind visualising and analysing the data was to do a seasonal analysis, but because December 2017 was included in the set, this is normally grouped with the summer of the next year (2018), due to this we decided to split each season only using 2 months.

- Months 1 & 2–Summer
- Months 3 & 4 –Autumn
- Months 6 & 7–Winter
- Months 9 & 10–Spring
- Months 5, 8, 11 & 12–Leftovers



# The Wrangling Process (Continued)

For every chunk in the CSV, process drops all rows with NaN / Null values in a column, converts Duration Seconds to a datetime object – which then gets converted to its necessary HH:MM:SS format, converts Arrival & Departure Time to datetime format and strips any necessary empty space from the relevant columns.

After each chunk has been cleaned, it is then appended to the cleaned.csv.

```
def extractMonth(date):  
    month,date,rest = date.split('-')  
    return int(date)
```

To do the wrangling process we created the extract.py program (which uses pandas and the system module). The main function of this program was the extractMonth() function which split the month from DepartureTime column of an individual row.

```
if __name__ == "__main__":  
    column_names = ["DeviceId", "ArrivalTime", "DepartureTime",  
                    "DurationSeconds", "StreetMarker", "Sign", "Area", "StreetId",  
                    "StreetName", "BetweenStreet1", "BetweenStreet2", "Side Of Street",  
                    "In Violation", "Vehicle Present", "month"]  
    summer = pd.DataFrame(columns = column_names)  
    autumn = pd.DataFrame(columns = column_names)  
    winter = pd.DataFrame(columns = column_names)  
    spring = pd.DataFrame(columns = column_names)  
    leftovers = pd.DataFrame(columns = column_names)  
    csvname = input('Enter CSV File: ')  
    chunk_no = 1  
    df = pd.read_csv(csvname, sep=',', skiprows=1, nrows=10000, low_memory=False)  
    for index, row in df.iterrows():  
        print('Starting Chunk Number: {}'.format(chunk_no))  
        df.loc[index, 'month'] = extractMonth(row['DepartureTime'])  
        chunk_no += 1
```

# The Wrangling Process (Continued)

The program creates 5 new dataframes – summer, autumn, winter, spring & leftovers – and assigns each row to the appropriate season based on the month extracted from `extractMonth()`. Upon completion it converts those dataframes to the necessary seasonal CSV.

```
for index, row in df.iterrows():

    if row['month'] == 1 or row['month'] == 2:
        summer = summer.append(row)
        print('Added: {} to summer dataframe'.format(row['month']))
    elif row['month'] == 3 or row['month'] == 4:
        autumn = autumn.append(row)
        print('Added: {} to autumn dataframe'.format(row['month']))
    elif row['month'] == 6 or row['month'] == 7:
        winter = winter.append(row)
        print('Added: {} to winter dataframe'.format(row['month']))
    elif row['month'] == 9 or row['month'] == 10:
        spring = spring.append(row)
        print('Added: {} to spring dataframe'.format(row['month']))
    elif row['month'] == 5 or row['month'] == 8 or row['month'] == 11 or row['month'] == 12:
        leftovers = leftovers.append(row)
        print('Added: {} to leftovers dataframe'.format(row['month']))

print('CSV Creatings...')
summer.to_csv('summer.csv', mode='a', index=False)
autumn.to_csv('autumn.csv', mode='a', index=False)
winter.to_csv('winter.csv', mode='a', index=False)
spring.to_csv('spring.csv', mode='a', index=False)
leftovers.to_csv('leftovers.csv', mode='a', index=False)
sys.exit()
```

# The Sampling Process

As mentioned above, our sampling idea is to compare the beginning, the middle and the end of the csv with each other to determine if there is any disparity with the data. To further sample the data we chose 4 main columns that we thought would show disparity (if any) in our scatter matrices – Duration Seconds, In Violation, Vehicle Present & Side of Street.

We decided on using the seaborn library for this as it provides more customisation and an easier-to-read scatter matrix for comparisons.

```
1 import pandas as pd
2 import seaborn as sns
3 df1 = pd.read_csv('espring.csv', sep=',', low_memory=False)
4 df2 = pd.read_csv('rspring.csv', sep=',', low_memory=False)
5 df3 = pd.read_csv('vspring.csv', sep=',', low_memory=False)

1 df1['DurationSeconds'] = df1['DurationSeconds'].map(lambda x: pd.to_timedelta(x).seconds)
2 df2['DurationSeconds'] = df2['DurationSeconds'].map(lambda x: pd.to_timedelta(x).seconds)
3 df3['DurationSeconds'] = df3['DurationSeconds'].map(lambda x: pd.to_timedelta(x).seconds)

1 df1['In Violation'] = df1['In Violation'].astype(int)
2 df1['Vehicle Present'] = df1['Vehicle Present'].astype(int)
3 df2['In Violation'] = df2['In Violation'].astype(int)
4 df2['Vehicle Present'] = df2['Vehicle Present'].astype(int)
5 df3['In Violation'] = df3['In Violation'].astype(int)
6 df3['Vehicle Present'] = df3['Vehicle Present'].astype(int)

1 sns.set(style="ticks", color_codes=True)
2 df1plot = sns.pairplot(df1, vars=['DurationSeconds', 'Vehicle Present', 'In Violation', 'Side Of Street'])
3 df1plot.savefig('spring1.png')
```

The process of creating the scatter matrices was quite simple, we would load each csv into Jupyter:

- epspring.csv is Elijah's spring CSV – which is the first 10,000 rows
- rspring.csv is Rohit's spring CSV – which is the middle 10,000 rows
- vspring.csv is Vinesh's spring CSV – which is the last 10,000 rows
- This continues for each seasonal CSV – except the leftovers CSV.

Convert DurationSeconds from HH:MM:SS back to its original format (just displaying the actual seconds) and convert the necessary Boolean fields (In Violation & Vehicle Present) to a numerical based field as scatter matrices can only operate on numerical fields. After that it was the simple process of assigned the style to the matrix, the colour codes and assigning the variables.

# The Sampling Process (Continued)

From here on we will see the scatter matrices and what we found for each season. Please remember that:

- Figure 1 represents the first 10,000 rows
- Figure 2 represents the middle 10,000 rows &
- Figure 3 represents the last 10,000 rows

## Summer Matrices

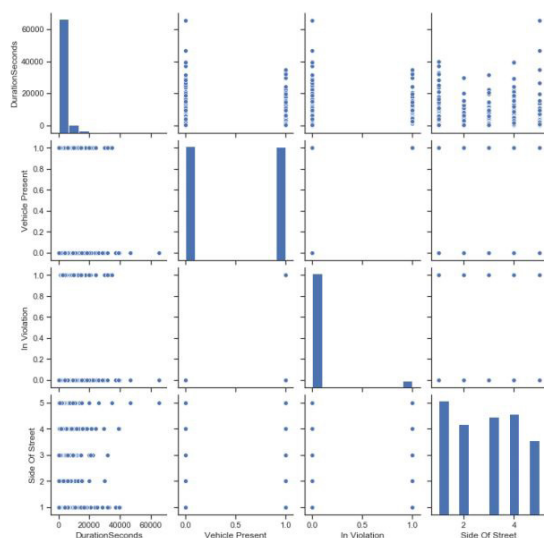


Figure 1

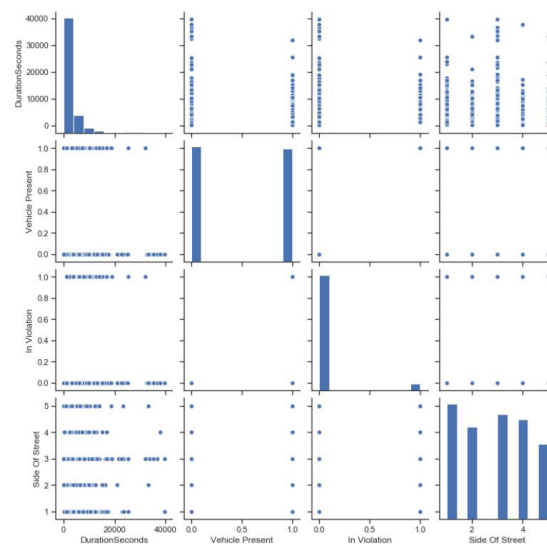


Figure 2

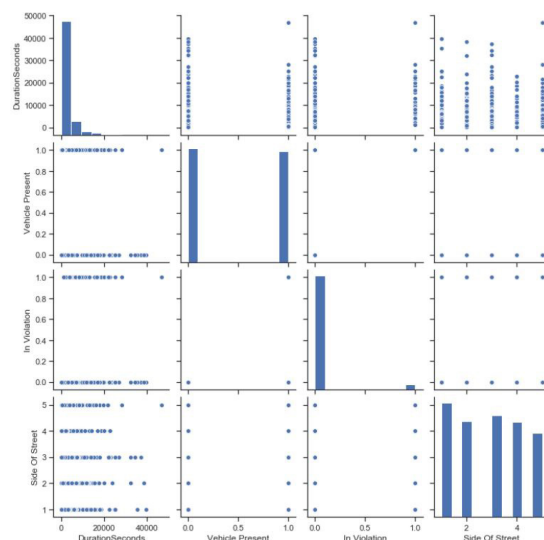


Figure 3

Comparing DurationSeconds with the other 3 columns; we can see that each of the figures show a disparity with the data, with more violations happening in the third sample set while the first sample set has less violations.

# The Sampling Process (Continued)

## Autumn Matrices

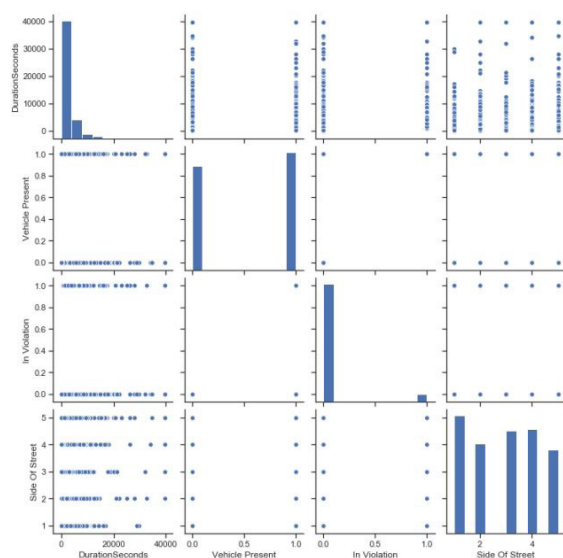


Figure 1

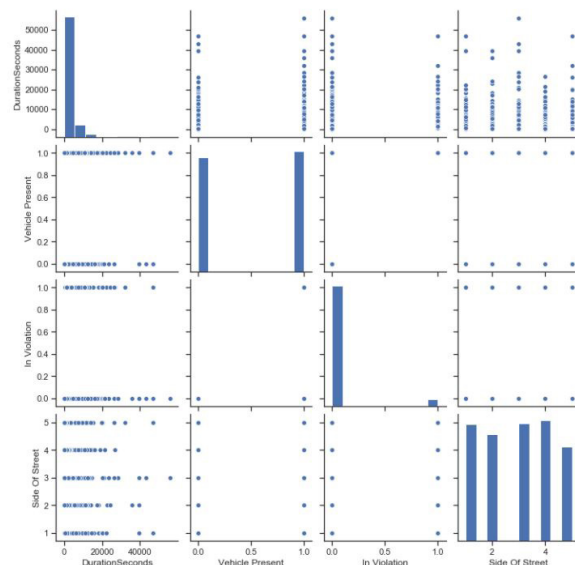


Figure 2

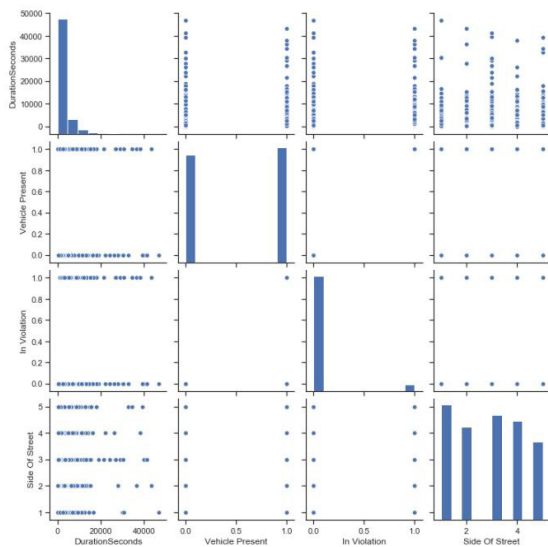


Figure 3

The Autumn sample data for each figure are showing more similarity between the 3 but there are small disparities between each of them.



# The Sampling Process (Continued)

## Winter Matrices

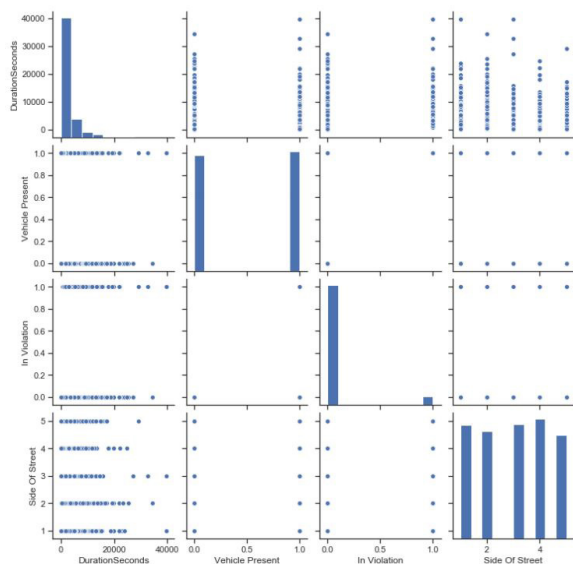


Figure 1

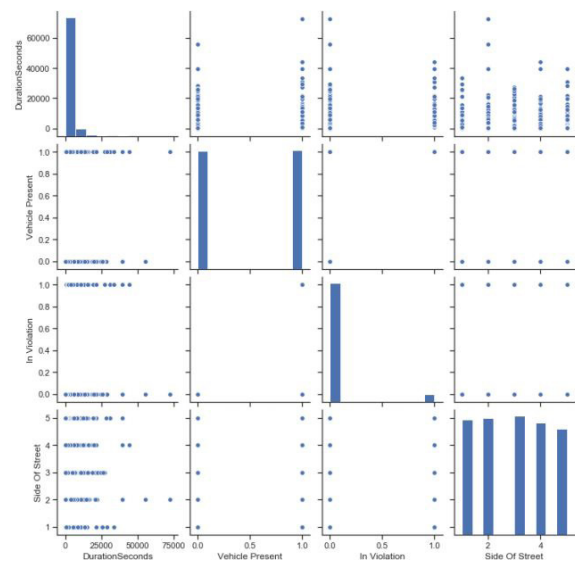


Figure 2

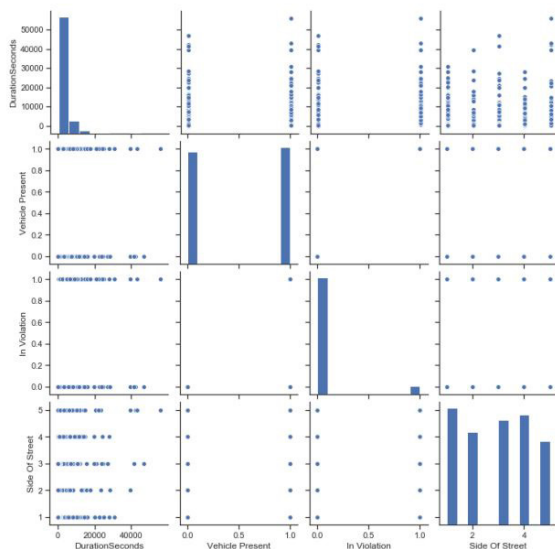


Figure 3

Winter is also showing a large disparity between the first and third sample set while the second sample set could be seen as a bridge between them – allowing for more of an accurate representation of Winter.

# The Sampling Process (Continued)

## Spring Matrices

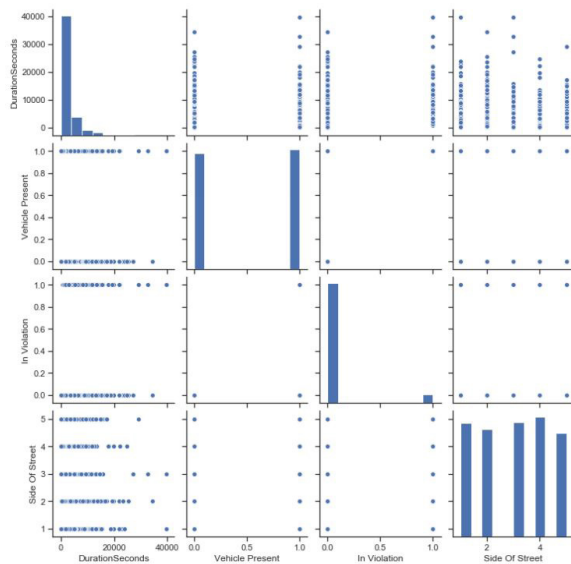


Figure 1

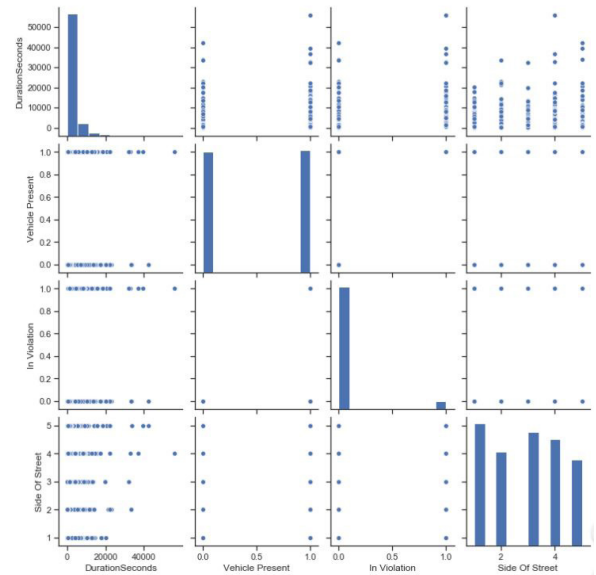


Figure 2

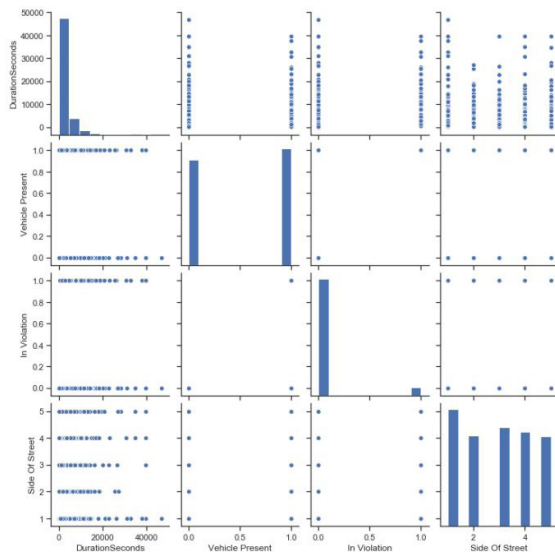


Figure 3

For Spring, The first and third sample set are showing to be quite similar, though the second sample set is drastically different then the two.

# Overview of Sampling & Road to Visualisation

- Comparing Duration Seconds with each of the columns (Side of Street, In Violation & Vehicle Present) for each sample set shows us different data plotted in each set
- Even though some sample sets show mild similarities, they aren't large enough for us to consider just dumping most of the sample set and only using 1 of those sets for our seasonal analysis
- Because of this it means we must combine them to give us accurate visualisations that we can use to conduct a proper analysis.

```
import pandas as pd
import sys

def main(csv1, csv2, csv3, newcsv, newercsv):
    column_names = ["DeviceId", "ArrivalTime", "DepartureTime",
                    "DurationSeconds", "StreetMarker", "Sign", "Area", "StreetId",
                    "StreetName", "BetweenStreet1", "BetweenStreet2", "Side Of Street",
                    "In Violation", "Vehicle Present", "month"]
    newcsv = pd.DataFrame(columns = column_names)

    df1 = pd.read_csv(csv1, sep=',', low_memory=False)
    df2 = pd.read_csv(csv2, sep=',', low_memory=False)
    df3 = pd.read_csv(csv3, sep=',', low_memory=False)

    print('Doing DF1')
    for i, r in df1.iterrows():
        newcsv = newcsv.append(r)

    print('Doing DF2')
    for ind, ro in df2.iterrows():
        newcsv = newcsv.append(ro)

    print('Doing DF3')
    for index, row in df3.iterrows():
        newcsv = newcsv.append(row)

    print('CSV Creatings...')
    newcsv.to_csv(newercsv, mode='a', index=False)

if __name__ == "__main__":
    csv1 = input('Enter 1 CSV File: ')
    csv2 = input('Enter 2 CSV File: ')
    csv3 = input('Enter 3 CSV File: ')
    newcsv = input('Enter New Name of Dataframe: ')
    newercsv = input('What is the name of the csv being created? ')
    main(csv1, csv2, csv3, newcsv, newercsv)
```

To merge the sample sets we created a program called **mergecsv.py** which takes in each individualsample set and the name of the csv that you wish to be merged into (in this case each season) and then loops through each individual dataframe and appends them to a dataframe, once this is complete it converts that dataframe to a CSV.



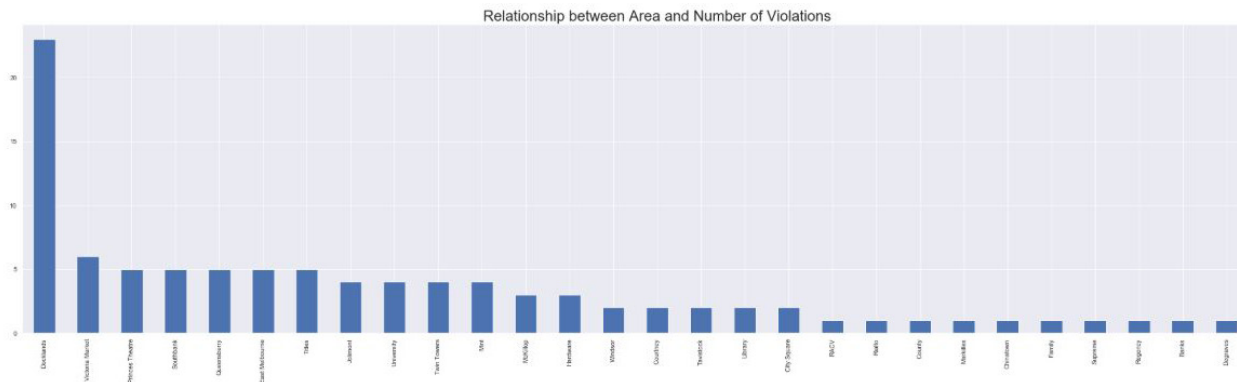
# Visualisation

A season file was assigned to each team member and we were tasked with coming up with as many visualisations as possible for the data. From these visualisations, we would pick out the most relevant ones and replicate them for the other seasons.

The teams knowledge of visualisation varied across the team, though everyone had to do some research and practice beforehand to refresh themselves with the technology or to learn the technology from scratch.

We used Anaconda Python with pandas, matplotlib, seaborn and numpy to generate multiple bar and pie charts.

Anaconda is widely used on Jupyter Notebooks to visualise data. Each different visualisation can be done within a cell. Jupyter Notebook helps with keeping the code look clean and organised. We found the ability to run specific cells instead of the whole program to be one of its most useful features.

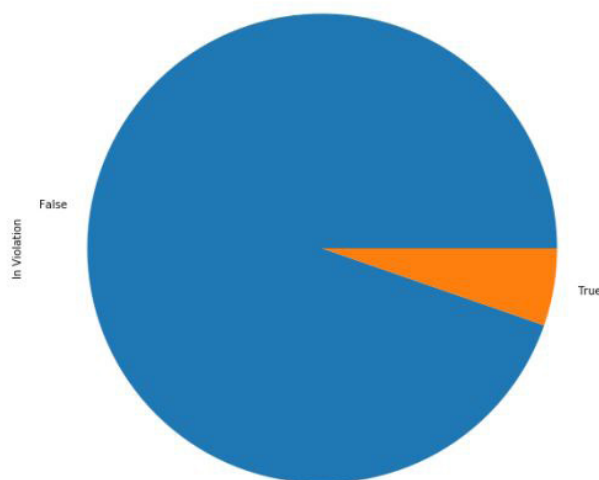


```
#Plot of Area and Average Duration
time = pd.DatetimeIndex(summer_data1['DurationSeconds'])
timenew = time.hour * 60 + time.minute
summer_data1['DS']=timenew

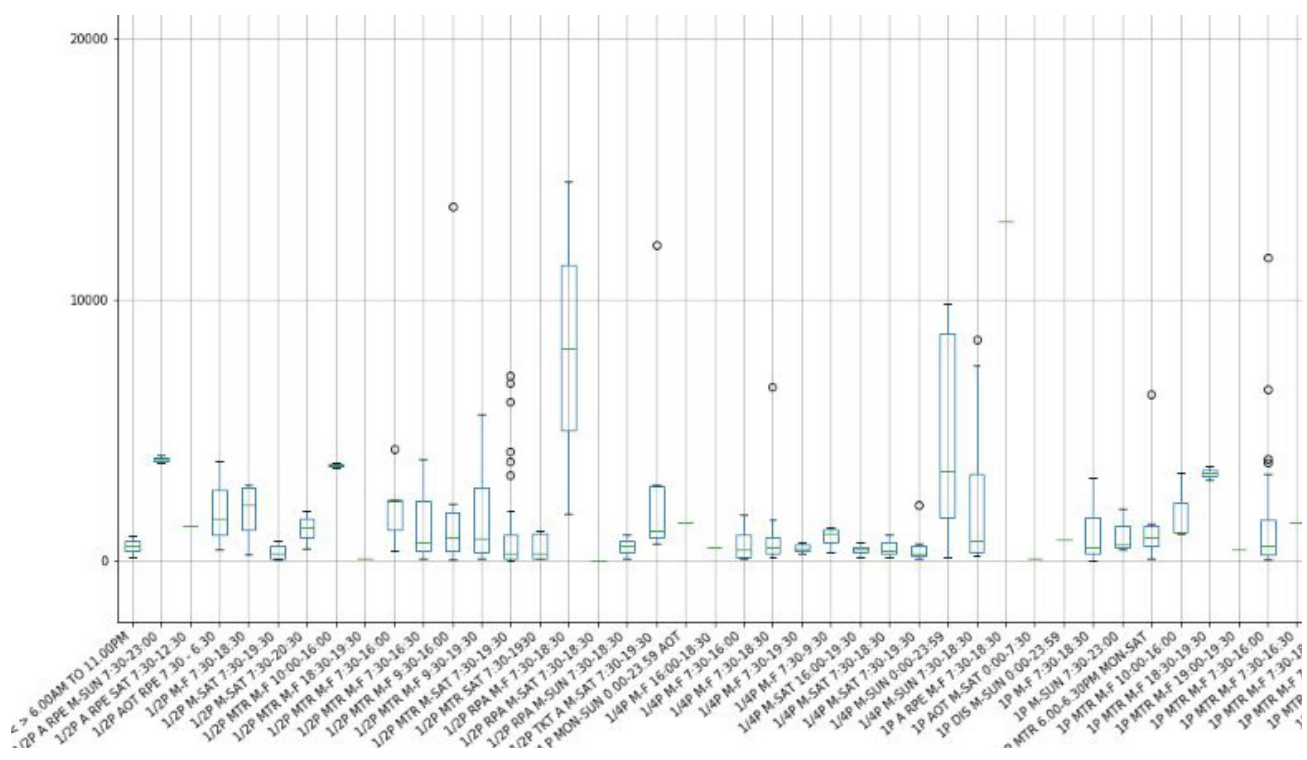
plt.figure(figsize=(40,6))
sns.boxplot(summer_data1['Area'],summer_data1['DS']);
plt.title('Relationship between Area and Duration',fontsize = 26);
```

# Visualisation (Continued)

Pandas is an easy to use data analysis and manipulation tool. Our graphs were generated with matplotlib, and seaborn was used for more in-depth graphs. Numpy was used to perform mathematical functions on the data to be used to generate graphs. Examples of the graphs obtained are:



A Pie Chart showing the ratio of parking violations in the summer



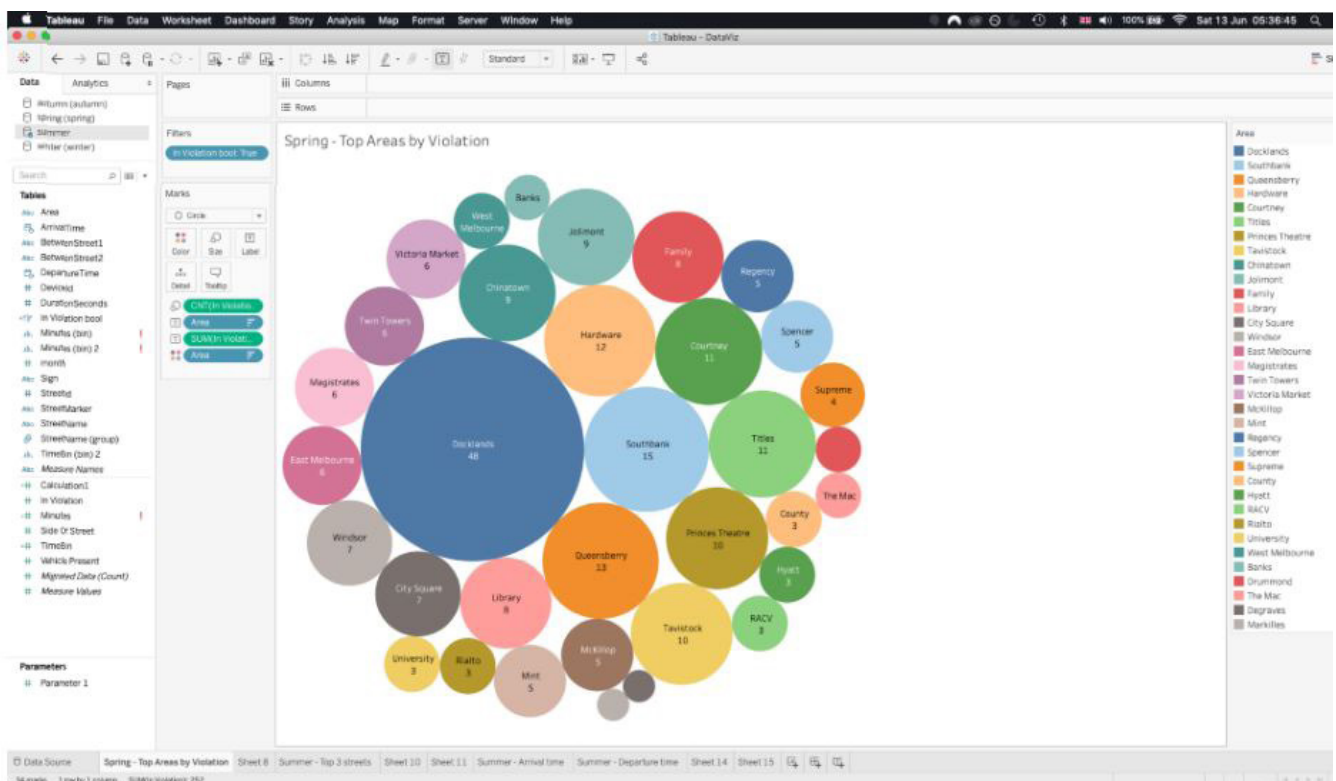
# Visualisation (Continued)

## The Switch to Tableau:

Our product sponsor then advised us to use the data visualisation software Tableau to generate charts. Tableau was relatively easy to learn and made the team more efficient. We were able to easily create pleasant looking and informative graphs at a much faster rate.

## What is Tableau:

Tableau is an intuitive data visualisation software that requires no coding or scripting to generate graphs. It automatically reads the dataset and correctly assigns a datatype to the columns. To generate a simple bar chart, one can simply drag and drop a column to each axis. We used our student emails to obtain a 1-Year license for the software.



# Visualisation (Continued)

## Using Tableau For the Visualisation

Our product sponsor then advised us to use the data visualisation software Tableau to generate charts. Tableau was relatively easy to learn and made the team more efficient. We were able to easily create pleasant looking and informative graphs at a much faster rate.

The task remained the same and the team made as many graphs as possible to understand the data we were working with even more. Tableau made the tasks more fun and easier than writing python code, and as a result we were more productive as a team, producing more visualisations in less time.

Out of all the visualisations generated, we focused on four key graphs that would help give a better overlook of the data we were working with. These graphs would also help us understand why some areas would be more popular than others, at what time most violations occur and possible explanations for those and much more.

The graphs we will focus on are:

- Top Areas by Violations
- Top 3 Streets in the top 3 Areas by violation and their signs
- The arrival day and time of the violations
- The departure day and time of the violations

## Top Areas by Violations

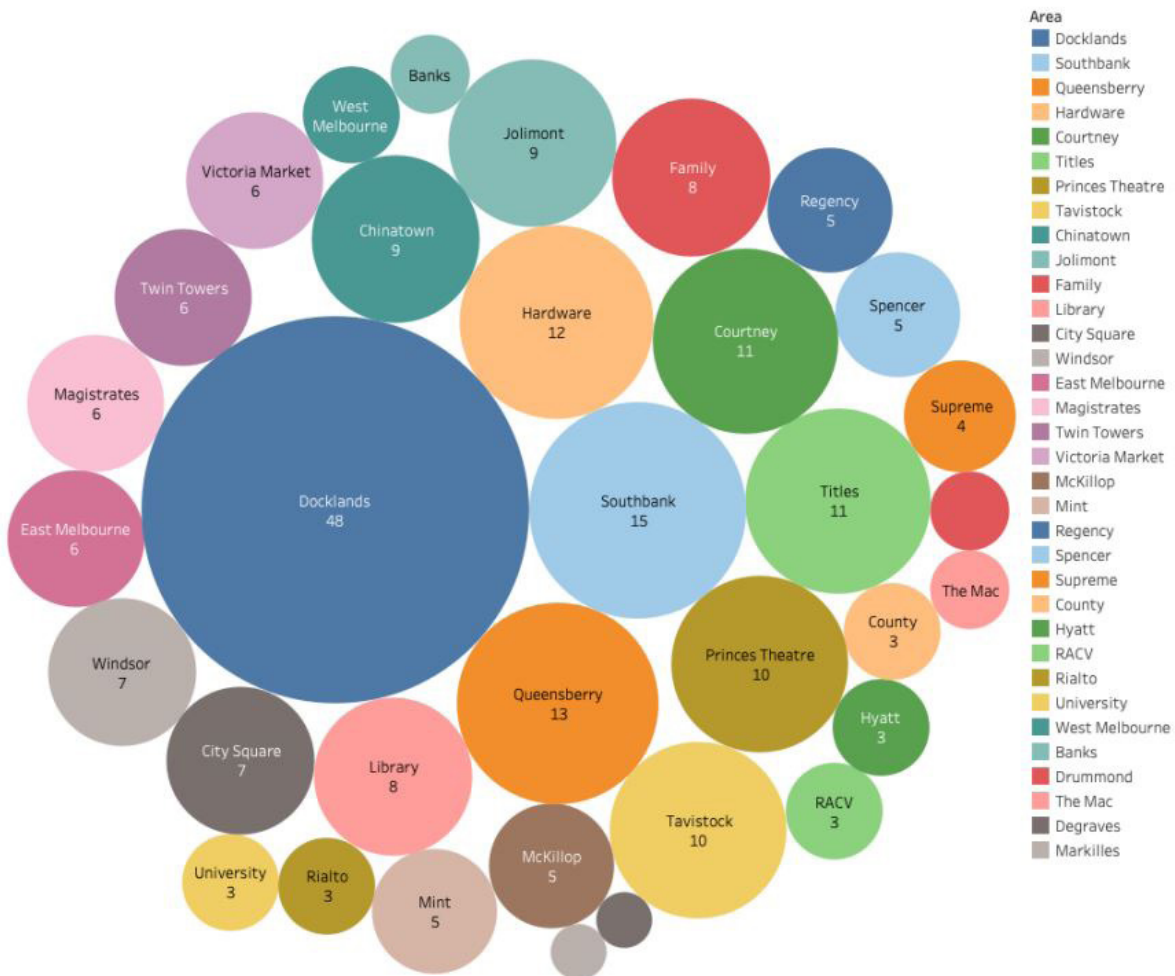
While having a general look at the data, we noticed that parking violations are not an issue throughout the CBD, but rather through some certain key areas in the CBD. These places had the most violations year-round. In order to find the cause of the high number of violations concentrated in a few areas and a solution to address the problem, we focused on the top 3 areas and analysed them.

# Visualisation (Continued)

## Summer:

Top 3 areas: Docklands, Southbank and Queensberry

Summer - Top Areas by Violation



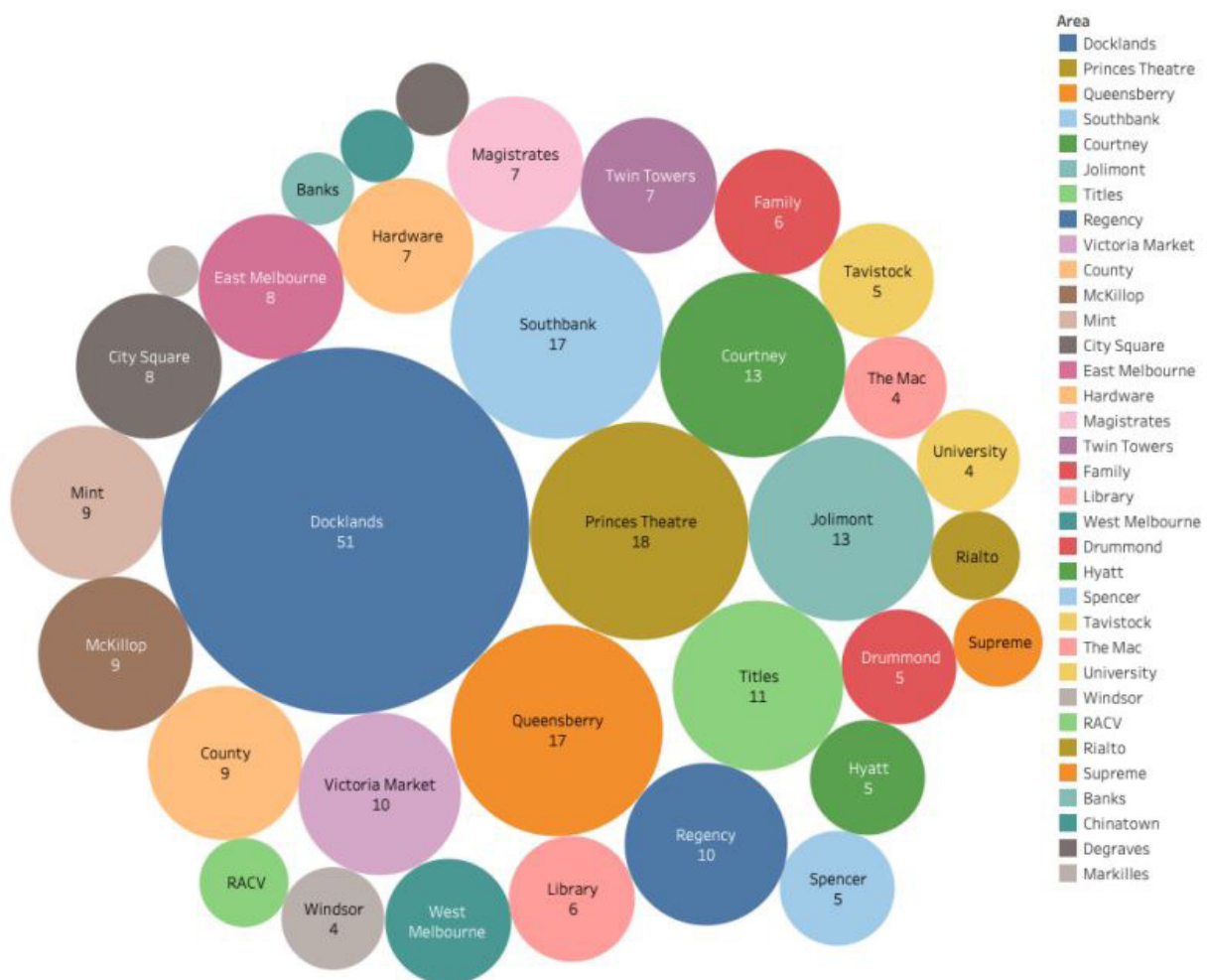
Area and sum of in violation. Color shows details about Area. Size shows count of in violation bool. The marks are labeled by Area and sum of in Violation. The data is filtered on in Violation bool, which keeps True.

# Visualisation (Continued)

## Autumn:

Top 3 areas: Docklands, Princess Theatre and Queensberry

Autumn - Top Areas by Violation



Area and sum of in violation. Color shows details about Area. Size shows count of in violation bool. The marks are labeled by Area and sum of in Violation. The data is filtered on in Violation bool, which keeps True.

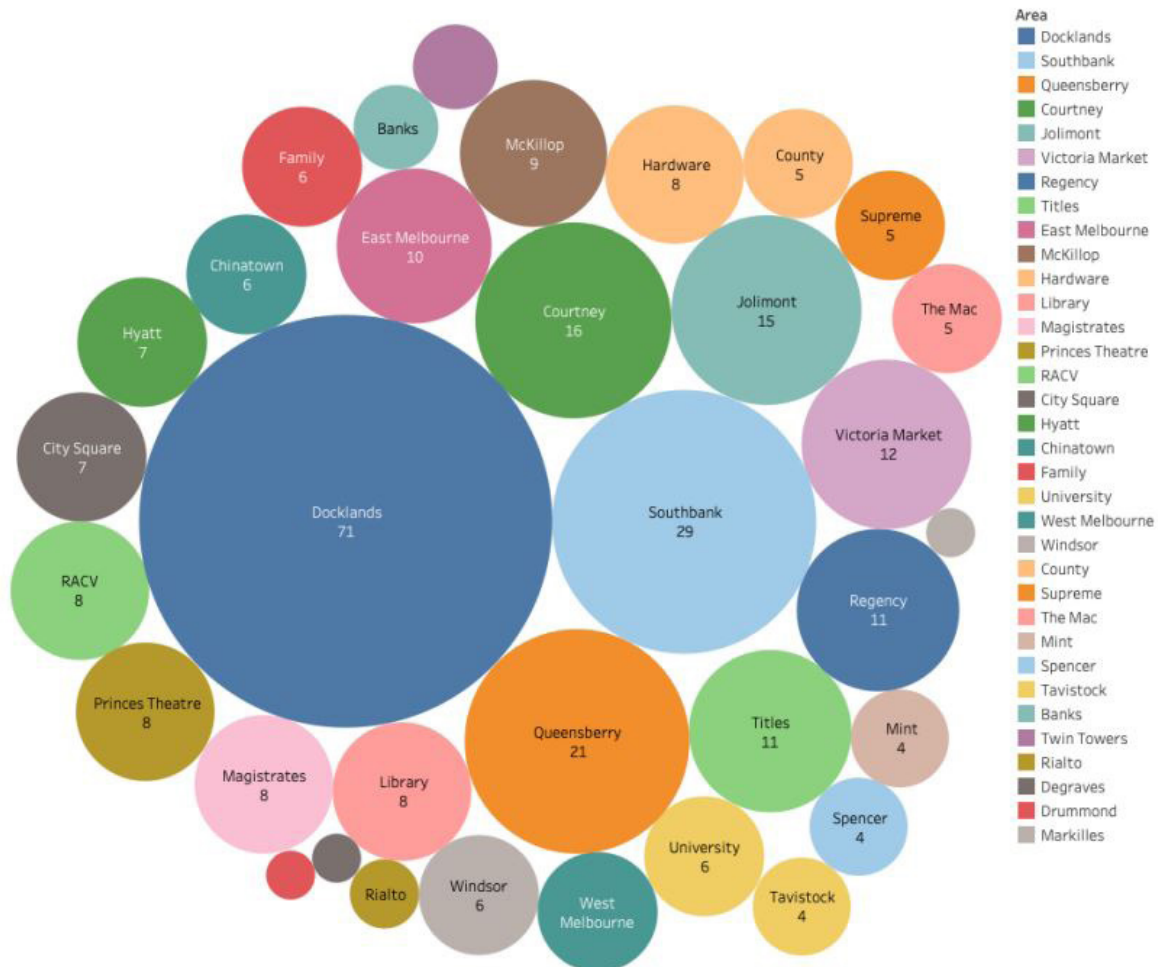


# Visualisation (Continued)

## Winter:

Top 3 areas: Docklands, Southbank and Queensberry

Winter - Top Areas by Violation



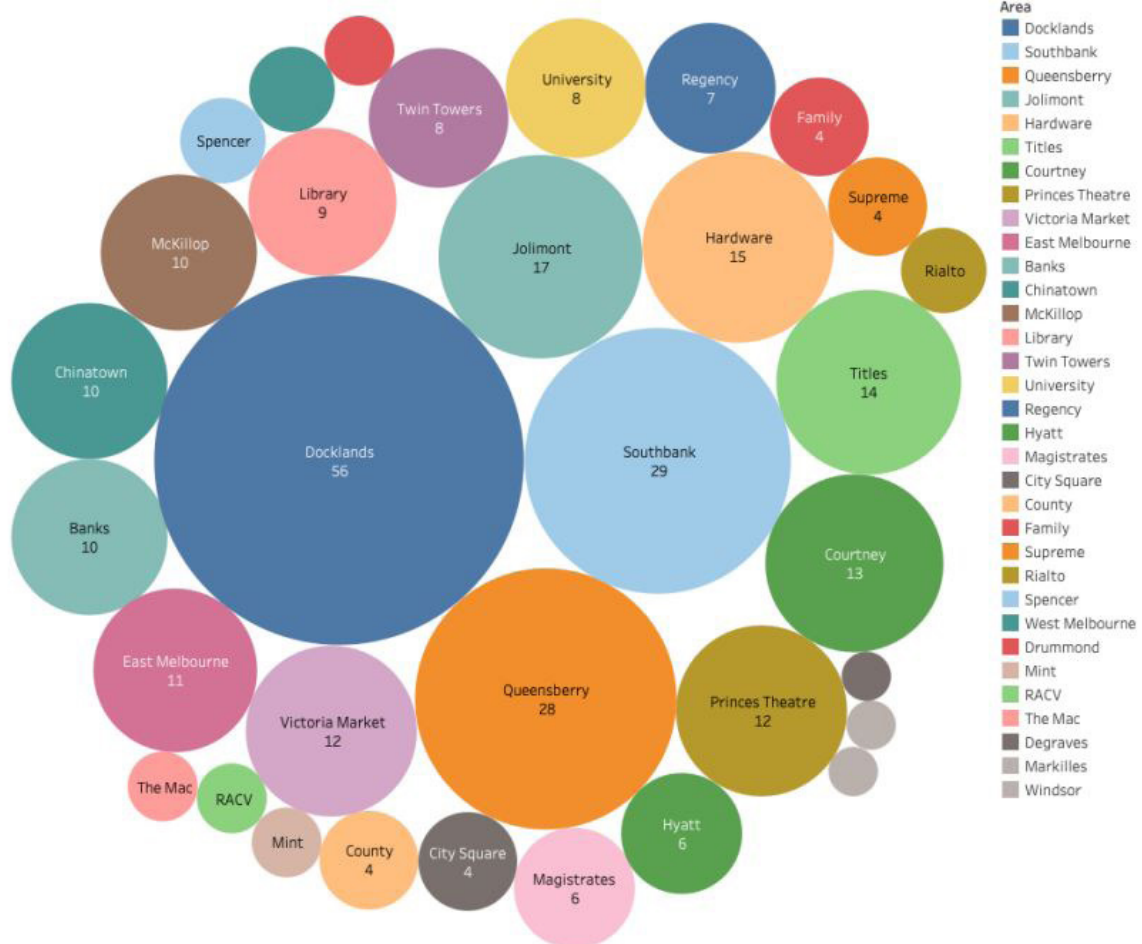
Area and sum of in violation. Color shows details about Area. Size shows count of in violation bool. The marks are labeled by Area and sum of in Violation. The data is filtered on in Violation bool, which keeps True.

# Visualisation (Continued)

## Spring:

Top 3 areas: Docklands, Southbank and Queensberry

Spring - Top Areas by Violation



Area and sum of in violation. Color shows details about Area. Size shows count of in violation bool. The marks are labeled by Area and sum of in Violation. The data is filtered on in Violation bool, which keeps True.



# Visualisation (Continued)

## Analysis for Top Areas:

The top area with the most violations is Docklands across all seasons. It exceeds the second top area by an average of 183%.

For summer, winter and spring, Southbank falls in second place and is closely followed by Queensberry.

For autumn, the top three differs a bit as the top three areas are: Docklands, Princess Theatre and Queensberry. Princess Theatre being at the second spot instead of Southbank may be because of the Autumn Blitz, which was construction for the new Metro tunnels.

It can also be noticed that the number of violations greatly increase in the winter. For example, docklands violations increased by 48%. This might be because people choose to drive to work rather than take public transport due to the cold temperatures in winter.

## Top 3 Areas in top 3 Areas by violation and their signs

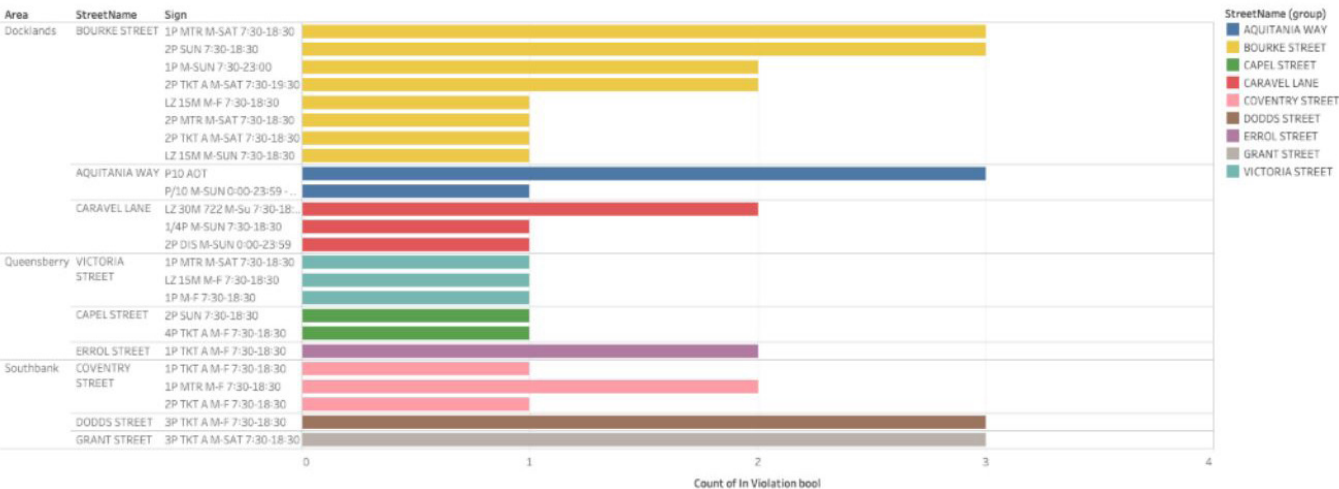
Now that the top 3 areas in terms of violations have been found, the team decided to dig deeper and see which streets in those areas have the most violations. We found out that it follows the same pattern as the top areas; it's only a few streets that make up for the majority of violations. For example, for Docklands, Bourke Street contributes for more than half of all the violations. As such, we decided to investigate the top 3 streets and the signs that were violated.

# Visualisation (Continued)

## Summer:

Bourke Street has the most signs violated compared to the other 8 streets. 8 different signs were violated on Bourke Street, while the other streets violate at most 3 signs. It is worth noting that the least amount of violated street signs has been in the summer.

Summer - Top 3 streets

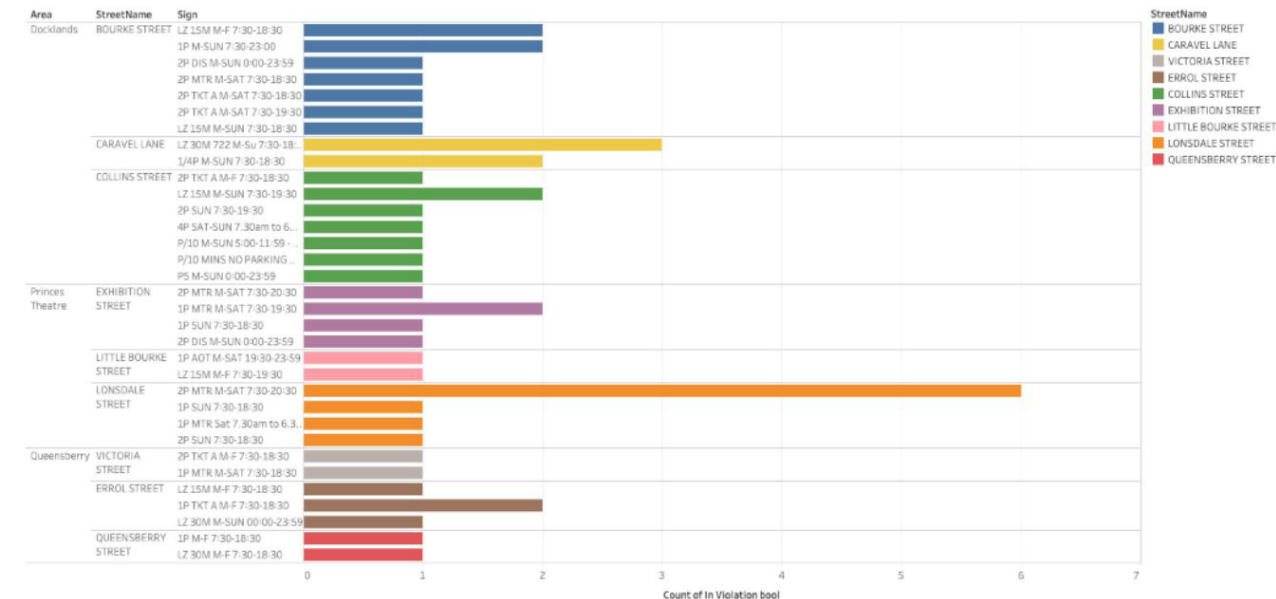


Count of In Violation bool for each Sign broken down by Area and StreetName. Color shows details about StreetName (group). The data is filtered on In Violation bool, which keeps True. The view is filtered on Area and StreetName. The Area filter keeps Docklands, Queensberry and Southbank. The StreetName filter keeps 9 members.

## Autumn:

Lonsdale Street has the most violations with the sign "2P MTR M-SAT 7:30-20:30" being the most frequently violated. Collins Street has the most violations on different signs. The violations seem to be minimal with signs being violated at most twice, except the extreme cases.

Autumn - Top 3 streets



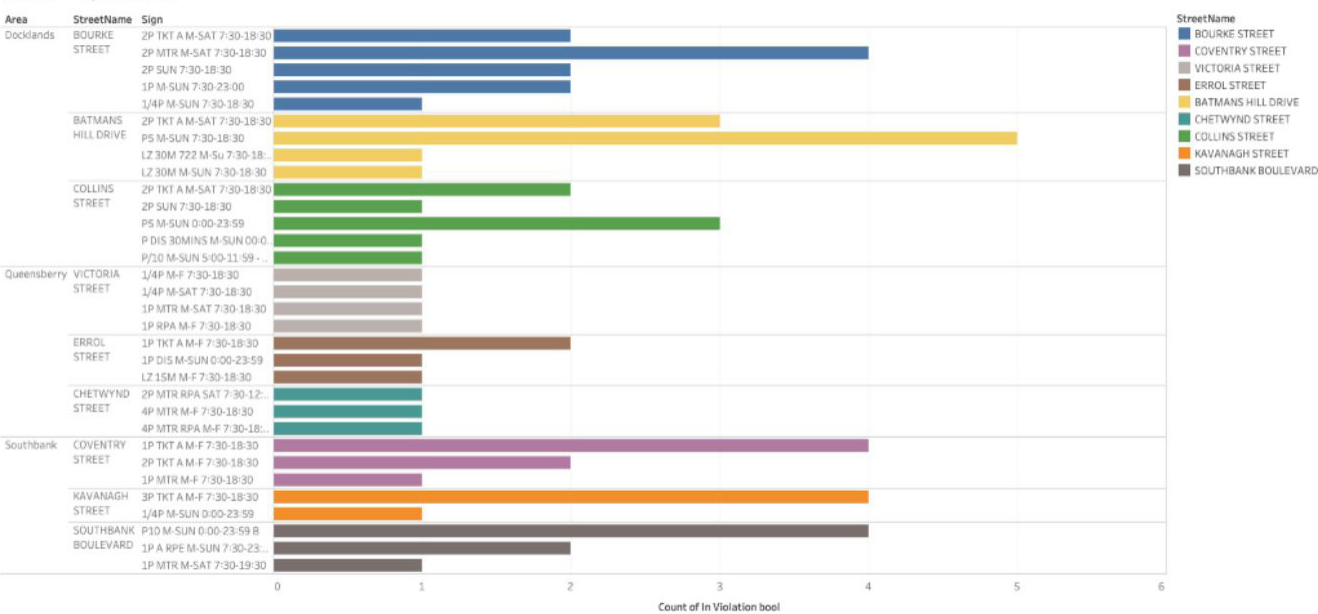
Count of In Violation bool for each Sign broken down by Area and StreetName. Color shows details about StreetName. The data is filtered on In Violation bool, which keeps True. The view is filtered on Area and StreetName. The Area filter keeps Docklands, Princes Theatre and Queensberry. The StreetName filter keeps 9 of 102 members.

# Visualisation (Continued)

## Winter:

Docklands and Southbank have an unusual spike in the number of violations on their top 3 roads. Queensberry seems mostly stable with the signs being violated only once, except on Errol Street

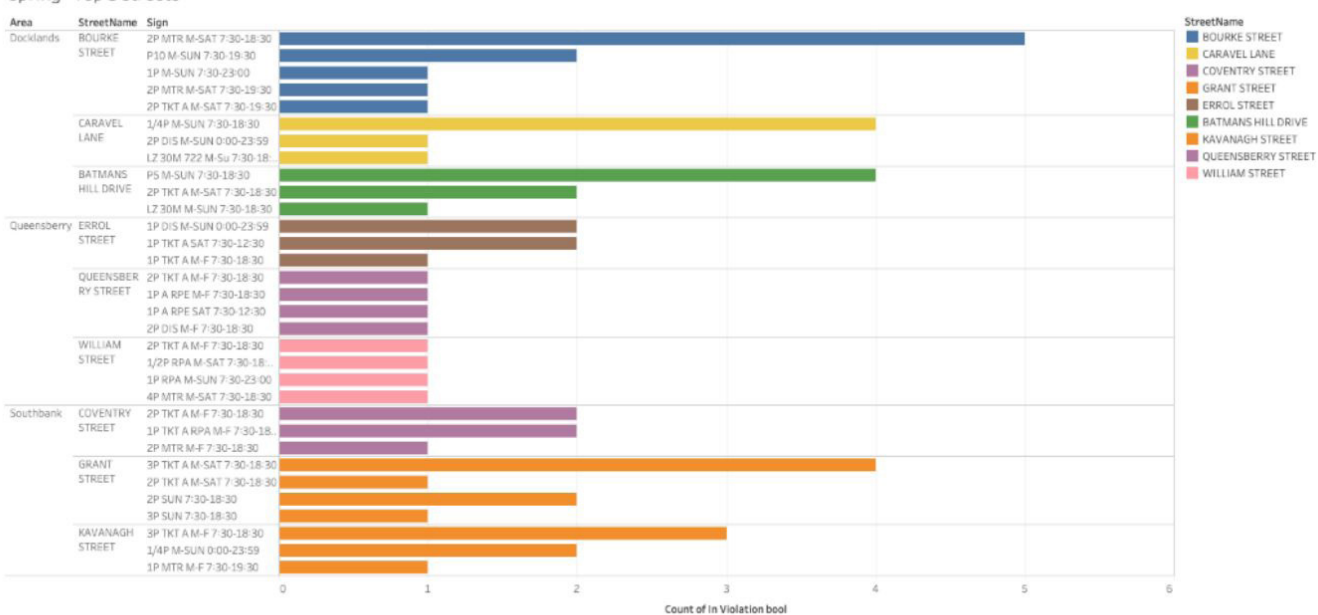
Winter - Top 3 streets



## Spring:

Caravel Lane, Batmans Hill Drive and Grant Street all seem to experience a spike in violations, while Bourke Street maintains a high number of violations. All streets violate between 2-5 signs.

Spring - Top 3 streets



Count of In Violation bool for each Sign broken down by Area and StreetName. Color shows details about StreetName. The data is filtered on In Violation bool, which keeps True. The view is filtered on Area and StreetName. The Area filter keeps Docklands, Queensberry and Southbank. The StreetName filter keeps 9 members.

# Visualisation (Continued)

## Analysis for top streets and signs:

When taking a general view of the data, it can be noticed that summer had the least signs violated (24) in the top three streets when compared to autumn (33), winter (32) and spring (32).

We can also notice that Bourke Street is consistently the top street in Docklands across all season. However, the other streets for the other seasons keep on changing.

The street signs seem to vary by seasons and not really have a pattern. The top sign changes every season, for example the top sign on Bourke Street for summer has not been violated in autumn where a sign was violated two times.

## Arrival day and time of the violations:

We wanted to understand why those signs were violated at this frequency.

The violations cannot be a mistake if multiple people violate the same sign.

Were the people arriving too early or late?

Did they park illegally on a loading zone?

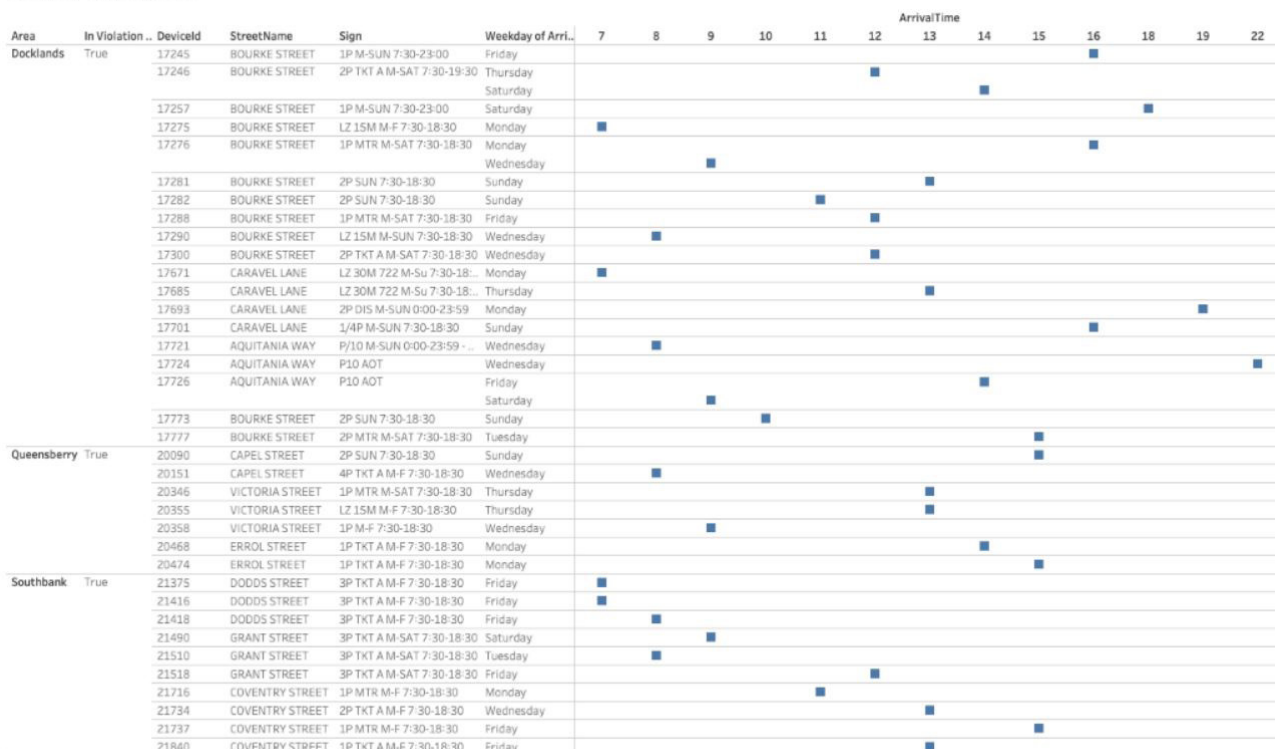
We plotted the signs obtained from the previous visualisations with the arrival day and time to help us answer these questions.

# Visualisation (Continued)

## Summer:

The arrival times of the violations can be clustered, and groups of people can be identified – We can see some people arrive early, probably to go to work and some people arrive midday.

Summer - Arrival time



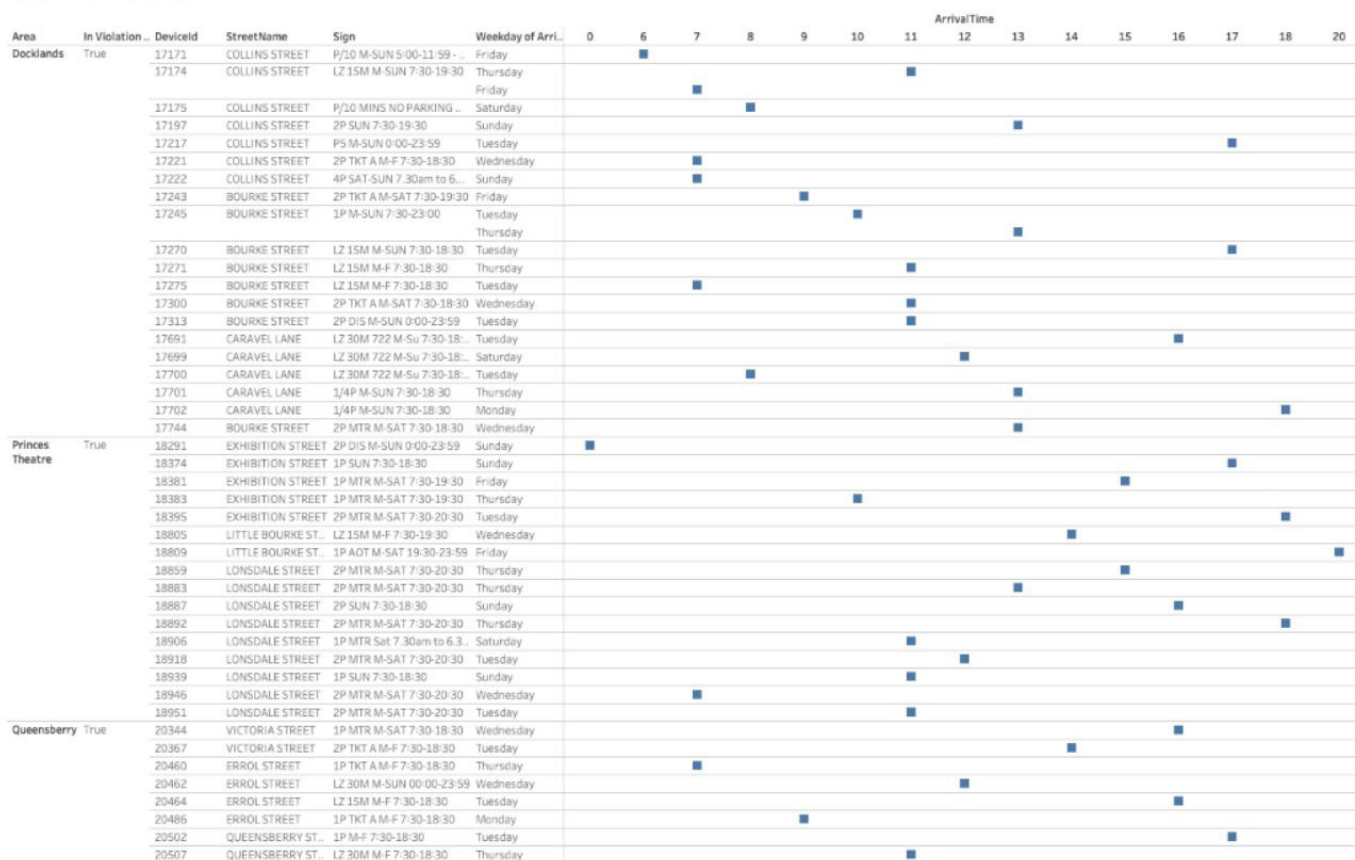
The view is broken down by ArrivalTime Hour vs. Area, In Violation bool, DeviceId, StreetName, Sign and ArrivalTime Weekday. The view is filtered on Area, In Violation bool and StreetName. The Area filter keeps Docklands, Queensberry and Southbank. The In Violation bool filter keeps True. The StreetName filter keeps 9 members.

# Visualisation (Continued)

## Autumn:

An interesting change in pattern can be observed here, as there does not seem to be that many early commuters in Autumn, but there is an increase in the amount of the people who arrive and park in the late morning and throughout the afternoon.

Autumn - Arrival time



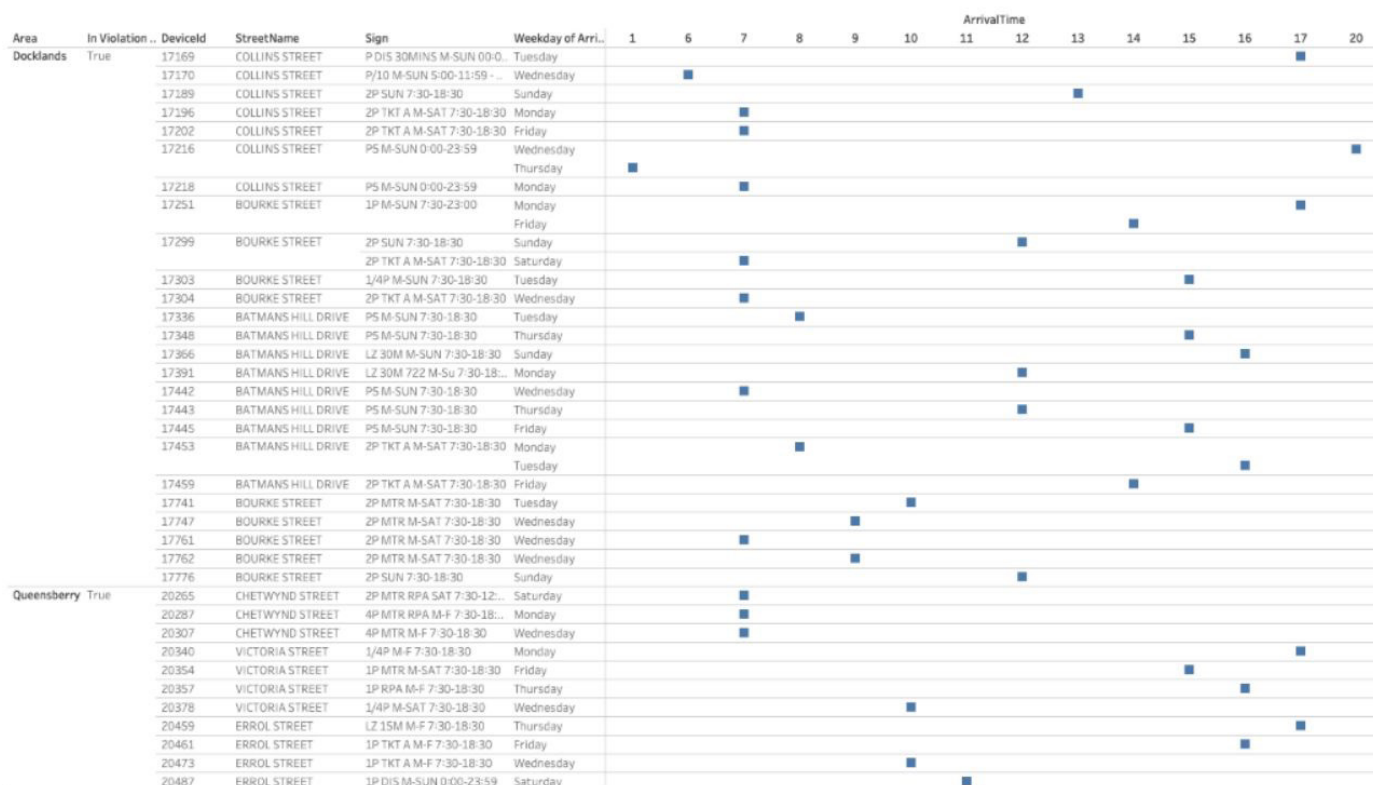
The view is broken down by ArrivalTime Hour vs. Area, In Violation bool, DeviceId, StreetName, Sign and ArrivalTime Weekday. The view is filtered on Area, In Violation bool and StreetName. The Area filter keeps Docklands, Princes Theatre and Queensberry. The In Violation bool filter keeps True. The StreetName filter keeps 9 members.

# Visualisation (Continued)

## Winter:

In winter, people can be clustered again. A high number of violations occur when people arrive between 07:00 and 08:00. There's no real cluster during the afternoon, but in the evening, violations start occurring again. late morning and throughout the afternoon.

Winter - Arrival time



The view is broken down by ArrivalTime Hour vs. Area, In Violation bool, DeviceId, StreetName, Sign and ArrivalTime Weekday. The view is filtered on Area, In Violation bool and StreetName. The Area filter keeps Docklands, Princes Theatre and Queensberry. The In Violation bool filter keeps True. The StreetName filter keeps 9 members.

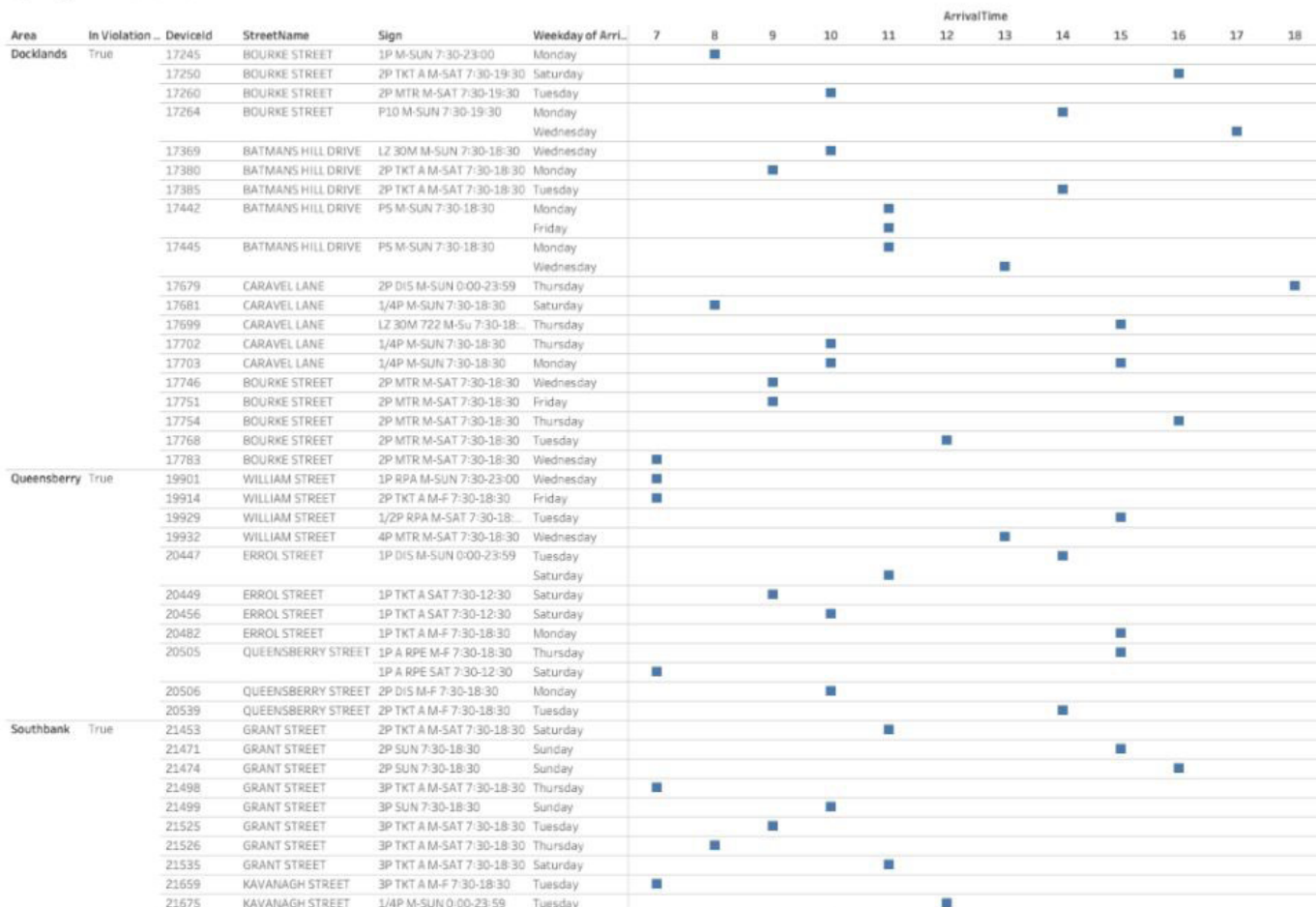


# Visualisation (Continued)

## Spring:

Clusters cannot be identified in spring, people seem to arrive consistently throughout the day.

Spring - Arrival time



## Analysis for arrival day and time

For **summer** we can group the arrival times of the violations in groups of 07:00–09:00 and 12:00–16:00. They can be grouped as people who are going to work and leaving work.

For **autumn**, we can notice that there are not many violations in the early morning but most of them occur in the mid-day at 11:00 until the evening at 17:00

For **winter**, a pattern can be noticed of arrivals at 07:00. This may be due to people choosing to drive to work in the winter, instead of waiting for public transport in the cold. The other times seem to be scattered and not much can be derived from them.

For **spring**, we cannot group the arrival times as they are evenly scattered across the day.



# Visualisation (Continued)

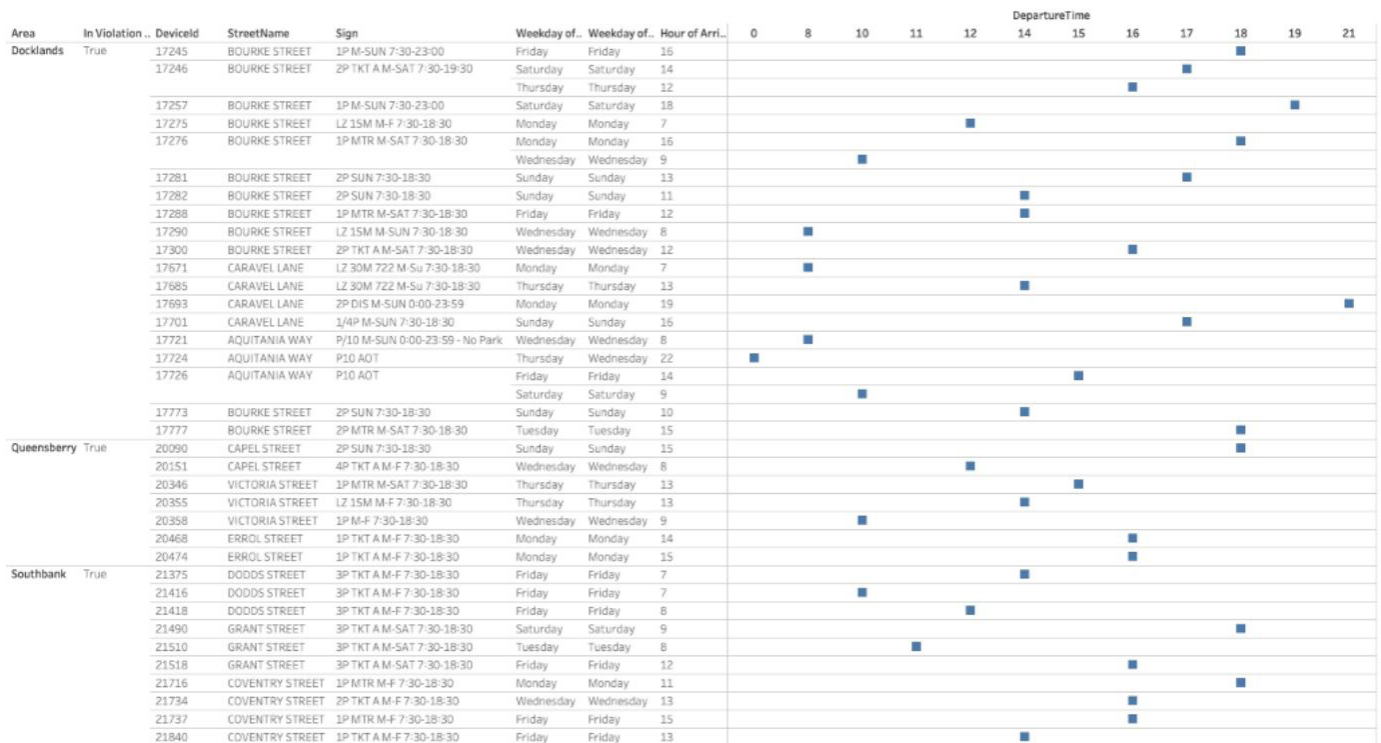
## Departure day and time of the violations:

For the following graphs, we decided to plot the departure time against the arrival time, so we would be able to get an idea of the length of stay of the vehicles. How long does the average person overstay their parking?

### Summer:

We noticed that for summer, the departure times of the violations occur in the evening, which indicates that people arrive early for work, park their cars all day long and leave in the evening. They tend to overstay for a few minutes only.

Summer - Departure time



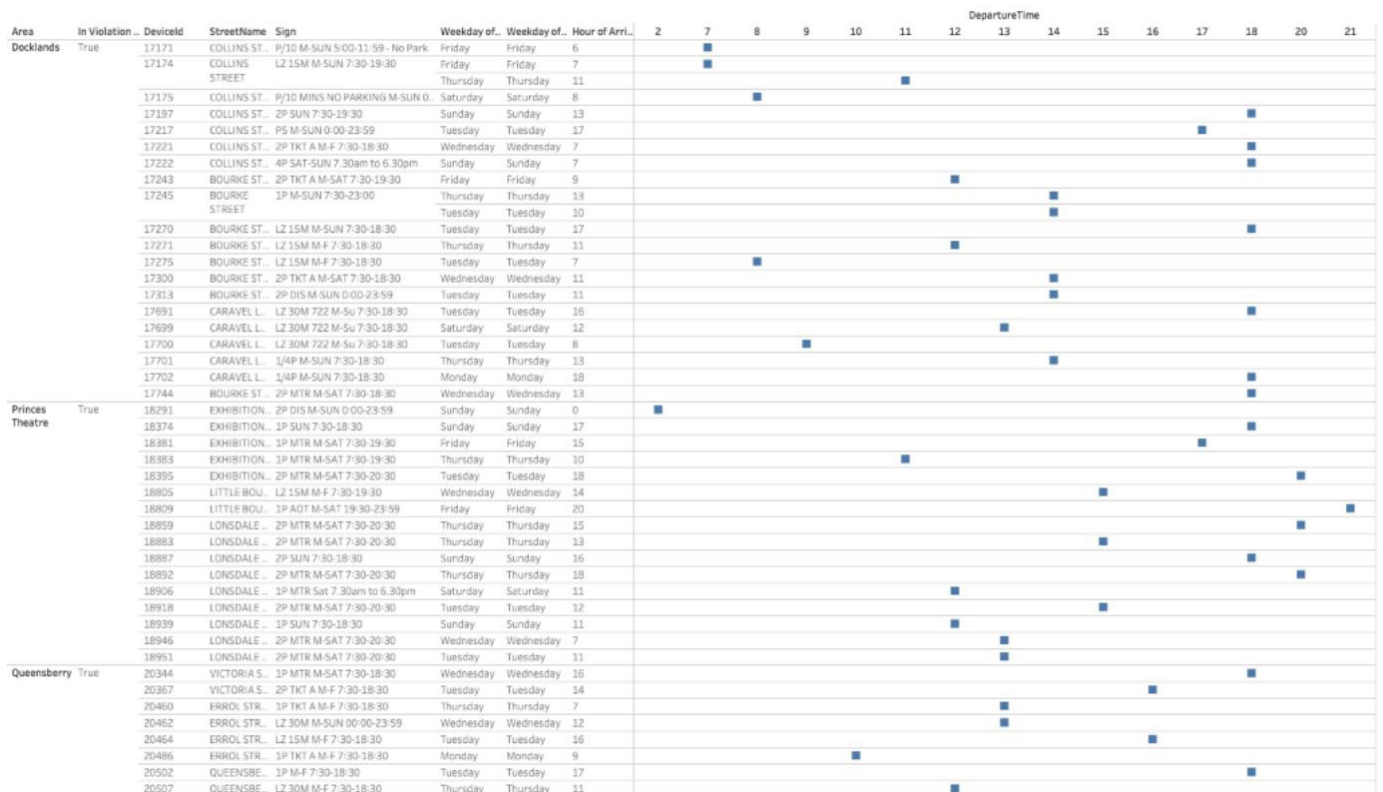
The view is broken down by DepartureTime Hour vs. Area, In Violation bool, DeviceId, StreetName, Sign, DepartureTime Weekday, ArrivalTime Weekday and ArrivalTime Hour. The view is filtered on Area, In Violation bool and StreetName. The Area filter keeps Docklands, Queensberry and Southbank. The In Violation bool filter keeps True. The StreetName filter keeps 9 members.

# Visualisation (Continued)

## Autumn:

The same pattern as summer can be observed, but in autumn, there are barely any violations from people who leave in the morning itself. Most of the violations are by people who left in the evening.

Autumn - Departure time



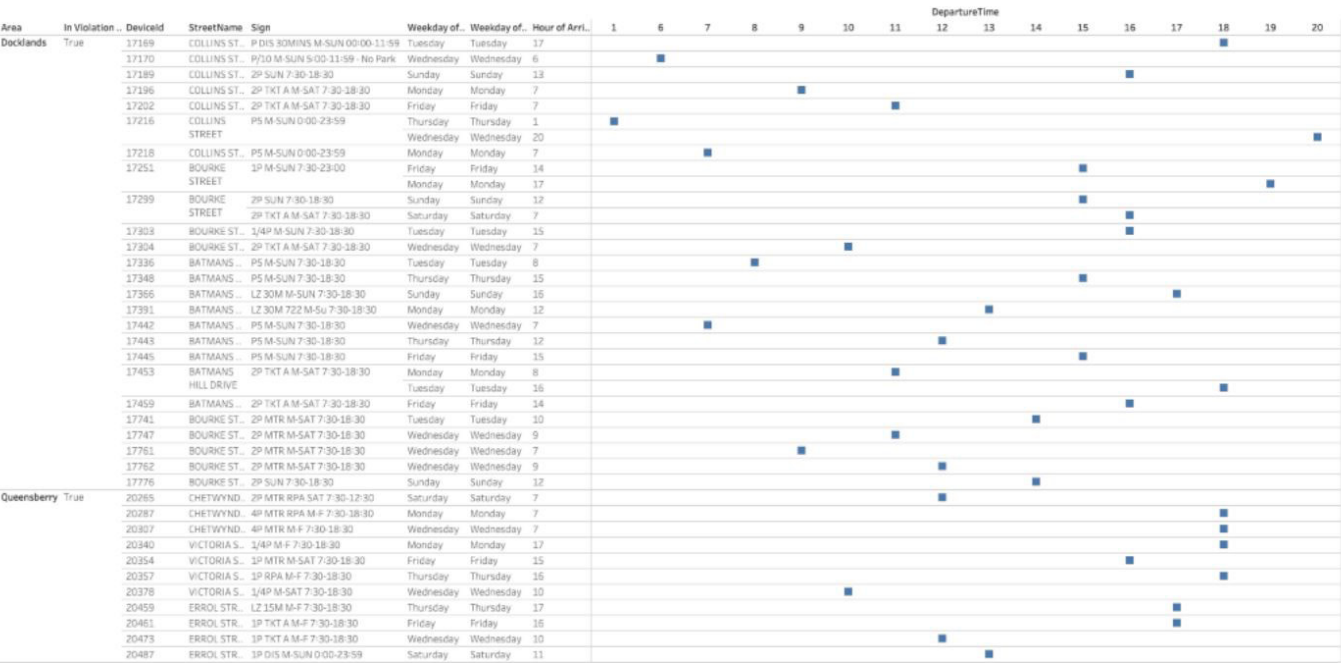
The view is broken down by DepartureTime Hour vs. Area, In Violation bool, DeviceId, StreetName, Sign, DepartureTime Weekday, ArrivalTime Weekday and ArrivalTime Hour. The view is filtered on Area, In Violation bool and StreetName. The Area filter keeps Docklands, Princes Theatre and Queensberry. The In Violation bool filter keeps True. The StreetName filter keeps 9 members.

# Visualisation (Continued)

## Winter:

The same pattern can be seen here, very few violations by the people who leave in the morning and the violations increase as the departure time in-creases

Winter - Departure time



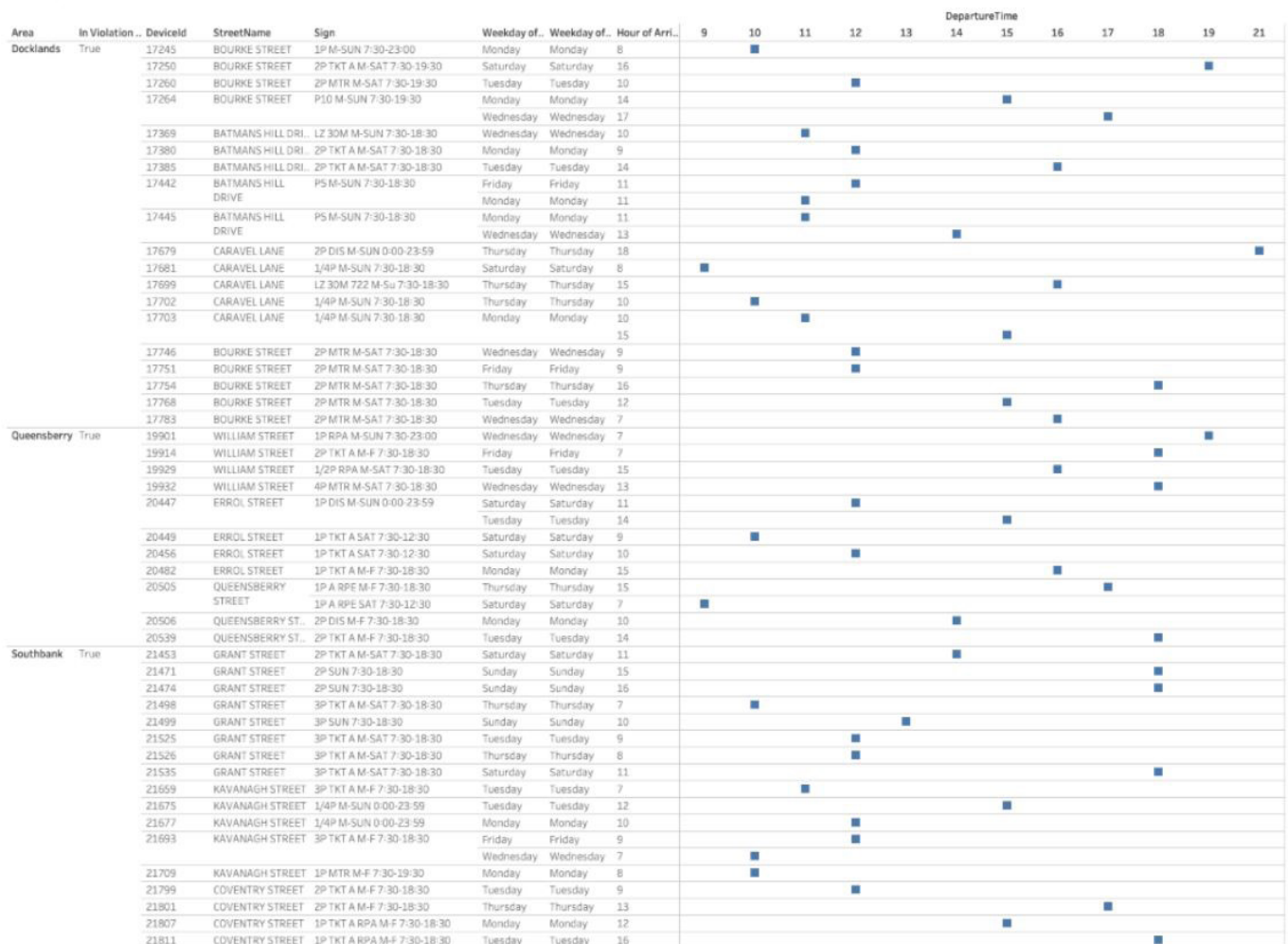
The view is broken down by DepartureTime Hour vs. Area. In Violation bool, DeviceId, StreetName, Sign, DepartureTime Weekday, ArrivalTime Weekday and ArrivalTime Hour. The view is filtered on Area, In Violation bool and StreetName. The Area filter keeps Docklands, Princes Theatre and Queensberry. The In Violation bool filter keeps True. The StreetName filter keeps 9 members.

# Visualisation (Continued)

## Spring:

The pattern breaks in spring, violations occur across all departure times. Early leavers seem to get as many violations as the rest.

Spring - Departure time



## Analysis for departure day and time:

Across all seasons, we noticed that the violations were either due to overstaying their parking by 30 minutes – 1 hour or by parking illegally at some spots, for example, on loading zones. The reason why most violations were due to overstaying a few minutes only might be because people are not willing to pay for an extra hour of parking only to be using it for a few minutes.

One such scenario could be someone overstaying their parking accidentally by bumping into a friend or grabbing a coffee on the way back to their car. They may also have some extra work that day, which makes them stay for a little longer than expected at work. Parking, especially for long hours can become expensive and paying for an extra hour is not worth it. People would rather take the chance and not get caught than pay.

# Conclusions

## How this data can help the City of Melbourne:

If the City of Melbourne wants to decrease the violations, they could extend each of the signs by an hour or two as that was the overage length of overstay on the parkings. Alternatively, if the city wants to make more money and profit on this situation, they could send out more officers during the timeslots mentioned above where we noticed an increase in arrival times.

To better cater to the needs of the citizens, instead of providing only hourly extensions on the parking, it might be a better idea to allow extensions by the quarter hour. People will be more willing to pay for the extra 15 minutes they will use of the parking, rather than pay for another hour but only use 15 minutes of that hour.

Additionally, a new charging scheme could be introduced. This would require better sensors to be used by the city as the new scheme would involve charging by the minute. If someone parks for 1 hour and 10 minutes, they will pay for 70 minutes instead of paying for 2 hours (120 minutes). This system, while being better for the citizens, will be expensive for the city as new sensors are and mobile apps may need to be developed. The city will also generate less revenue from fines as this scheme will greatly reduce the number of violations.

# References

## Reference List:

<https://data.melbourne.vic.gov.au/Transport/On-street-Car-Parking-Sensor-Data2017/u9sa-j86i>

<https://www.codetwo.com/admins-blog/how-to-split-csv-file-into-multiple-files-usingpowershell/>